

# Geniş Ölçekli Testlerde Örneklem Ağırlıklarının ve Olası Değerlerin Önemi

Serkan ARIKAN \* Ferah ÖZER \*\* Vuslat ŞEKER \*\*\* Güneş ERTAŞ \*\*\*\*

## Öz

Uluslararası Öğrenci Değerlendirme Programı (PISA), Yetişkin Becerileri Araştırması (PIAAC), Uluslararası Matematik ve Fen Bilimleri Eğilimi Araştırması (TIMSS) gibi geniş ölçekli testler, birincil amacı olan ölçme, değerlendirme ve gelişimi izlemenin yanı sıra, ülkelerin eğitim politikalarını belirleyici anahtar bir rol de üstlenmektedir. Makro boyutta politika değişiklikleriyle milyonlarca paydaşı etkileme potansiyeline sahip olan geniş ölçekli test verilerinin bilimsel olarak doğru istatistiksel yöntem ve tekniklerle analiz edilmesi büyük önem arz etmektedir. Çok sayıda değişken içeren ve kendine has özellikleri olan bu verilerin analizi uzmanlık gerektirmektedir. Bu araştırma geniş ölçekli test verilerinin analizinde dikkat edilmesi gereken noktaları açıklayarak ilgili analiz kodlarına yer vermektedir. Türkiye’de bu verilerin nasıl analiz edileceği konusunda sistematik olarak açılan ders neredeyse yoktur. Sınırlı sayıda yapılan çalıştayların da ulaşabildiği araştırmacı sayısı azdır. Bu sınırlılıkları aşabilmek için bu çalışmanın amacı hem olası değerler ve örneklem ağırlıkları hakkında farkındalık yaratarak araştırmacıları doğru analiz yöntemlerini araştırmaya ve kullanmaya yönlendirmek, hem de temel analizleri örneklendirerek araştırmacılara yol göstermektir. Elde edilen sonuçlar göstermektedir ki, örneklem ağırlıkları ve olası değerler doğru yöntemlerle kullanılmadığında hatalı sonuçlar elde edilebilmektedir. Standart hataların farklı kestirilmesinden kaynaklı olarak anlamlılık değerleri etkilenmektedir. Bu çalışmada IDB Analyzer kullanılarak t-testi ve çoklu regresyon, Mplus kullanılarak çok düzeyli regresyon analizi yapılmıştır.

**Anahtar Kelimeler:** Örneklem ağırlıkları, olası değerler, geniş ölçekli testler, IDB Analyzer, Mplus

## GİRİŞ

Uluslararası Öğrenci Değerlendirme Programı-PISA (The Program for International Student Assessment), Yetişkin Becerileri Araştırması-PIAAC (The Programme for the International Assessment of Adult Competencies), Uluslararası Matematik ve Fen Bilimleri Eğilimi Araştırması-TIMSS (Trends in International Mathematics Science Study) gibi geniş ölçekli testler, birincil amacı olan ölçme, değerlendirme ve gelişimi izlemenin yanı sıra, ülkelerin eğitim politikalarını belirleyici anahtar bir rol de üstlenmektedir (Bialecki, Jakubowski, & Wisniewski, 2017; Figazzolo 2009; Novoa & Yariv-Mashal, 2003; Steiner-Khamsi & Waldow, 2018). Geniş ölçekli testlerin tarihsel gelişimi incelendiğinde, ilk dönemlerde testleri geliştiren kuruluşların ülkeler arası karşılaştırma ve üst-alt başarı grupları oluşturma amaçlarını reddettiği, daha ziyade genel bir izlemeyi amaçladıkları görülmektedir (Landahl, 2018). İlerleyen dönemlerde ise bu testlere gerek yerel gerekse uluslararası medyanın ilgisinin artması ve çeşitli politik unsurların gereği olarak ülkelerarası karşılaştırma, ekonomik gösterege olarak kullanıma ve politikalar için gerekçe oluşturma gibi anlamlar yüklediği de görülmektedir. Addey, Sellar, Steiner-Khamsi, Lingard ve Verger (2017) ise ülkelerin bu tür testlere katılımlarının temel nedenini, politikalar için veri temelli bilgi sağlama, teknik kapasite-altyapı

\* Dr. Öğretim Üyesi, Boğaziçi Üniversitesi, Eğitim Fakültesi, İstanbul, serkan.arikan1@boun.edu.tr, ORCID ID: 0000-0001-9610-5496.

\*\*Araştırma Görevlisi, Boğaziçi Üniversitesi, Eğitim Fakültesi, İstanbul, e-mail: ferah.ozler@boun.edu.tr, ORCID ID: 0000-0001-8621-3522.

\*\*\* Araştırma Görevlisi, Boğaziçi Üniversitesi, Eğitim Fakültesi, İstanbul, e-mail: vuslat.seker@boun.edu.tr, ORCID ID: 0000-0002-3279-5544.

\*\*\*\* Araştırma Görevlisi, Boğaziçi Üniversitesi, Eğitim Fakültesi, İstanbul, e-mail: gunes.ertas@boun.edu.tr, ORCID ID: 0000-0001-8785-7768.

To cite this article:

Arıkan, S., Özer, F., Şeker, V. & Ertaş, G. (2020). The Importance of Sample Weights and Plausible Values in Large- Scale Assessments. *Journal of Measurement and Evaluation in Education and Psychology*, 11(1), 43-60. doi: 10.21031/epod.602765

Received: 06.08.2019

Accepted: 14.02.2020

geliştirme, maddi destek ve yardım yapma, uluslararası ilişkilerde öne çıkma, iç politikada karar alma, ekonomik gerekçeler, müfredat ve öğretim yaklaşımları için reformlar olduğunu vurgulamaktadır. Ayrıca Ekonomik İşbirliği ve Kalkınma Teşkilatı (OECD - The Organisation for Economic Co-operation and Development), UNESCO, Dünya Bankası gibi uluslararası kuruluşların, ülkelerin gerçekleştirdiği eğitim reformlarını izleme, yeni uygulamalarını anlamlandırmada ve gelişen ülkelere yapılacak hibe yardımlarına karar vermede bu verilerden yararlandığı da bilinmektedir (Addey & Sellar, 2018; Aydın, Selvitopu, & Kaya, 2018). Özetle, geniş ölçekli test verileri, ülkelere eğitim sistemlerini oluşturan elementlerin etkililiği, mevcut öğrenci, öğretmen, yönetici portföyleri ve eğitim sistemleri hakkında detaylı ve karşılaştırılabilir bilgi sağlamaktadır.

Addey vd. (2017)'nin de vurguladığı birincil amaca ilişkin olarak, son yıllarda dünya genelinde özellikle PISA, PIAAC ve TIMSS gibi geniş ölçekli testlerden elde edilen veriye dayalı bilgilerle, ülkelerin eğitim politikalarında makro ve mikro boyutlarda çeşitli reformlar gerçekleştirdiği bilinmektedir. Bu reformlar makro boyutta müfredat değişiklikleri, öğretim ve öğrenme süreçleri yaklaşımlarındaki değişimler, ölçme süreçlerindeki değişiklikler, dezavantajlı grupların oryantasyonu; mikro boyutlarda ise ders kitapları, materyaller, donanım-yazılımlar ve yerel okul kültürlerinde yapılan değişiklikler olarak sıralanabilir. Özellikle Fransa (Carvalho & Costa, 2015; Michel, 2017), Portekiz (Carvalho & Costa, 2015), Polonya (Bialecki et al., 2017), Macaristan (Carvalho & Costa, 2015), Almanya (Ertl, 2006), İsveç (Landahl, 2018), İsrail (Pizmony-Levy, 2018), İspanya (Tiana Ferrer, 2017) gibi birçok ülkenin müfredat reformları ya da köklü eğitim politikalarını meşrulaştırmada PISA ve TIMSS sonuçlarından faydalandığı bilinmektedir (Wiseman, 2013). Türkiye'de de özellikle 2000'li yılların başından itibaren gerçekleştirilen müfredat reformları ve ulusal sınav sistemlerinde yapılan değişiklikler bu makro girişimlere örnek olarak gösterilebilir. Özellikle 2013 ve 2018 yılında tüm seviye müfredatlarında gerçekleştirilen değişikliklerde PISA ve TIMSS ile ölçülen düşünme süreçlerine paralel olarak öğrencilerin akıl yürütme ve değerlendirme gibi becerilerini geliştirmelerini destekleyici öğrenme ortamlarının sağlanması gerekliliği sıkça vurgulanmaktadır. Türkiye'de yapılan merkezi sınavlar da bu değişimden etkilenmişlerdir. Ortaöğretime geçiş için kullanılan Liselere Giriş Sınavı (LGS) 2018 yılında yapılan değişiklikte birlikte üst düzey düşünme becerilerini ölçmeyi amaçlayan bir yapıya dönüştürülmüştür (MEB, 2018). Ayrıca ilk uygulaması 2015 yılında yapılan PISA benzeri eleştirel düşünme, problem çözme ve yorumlama gibi üst düzey becerileri ölçen ulusal ABİDE (Akademik Becerilerin İzlenmesi ve Değerlendirilmesi) programı da bu alandaki olumlu reformist girişimlere örnek olarak gösterilebilir.

Ayrıca son yıllarda geniş ölçekli testlerin sonuçlarının uluslararası, ulusal ve yerel medyada sıkça yer bulması (Martens & Niemann, 2010), PISA'ya ait özel uluslararası ve yerel bir medya algısının oluşması (Michel, 2017) sınavların sonuçlarına ilişkin ilgi ve farkındalığı da arttıran önemli bir faktör olarak öne çıkmaktadır (Froese-Germain, 2010; Gür, Çelik, & Özoğlu, 2012; Steiner-Khamsi & Waldow, 2018). Bu duruma paralel olarak eğitimciler ve birçok üniversite ve/veya araştırma merkezleri bünyesinde çalışan akademisyenler de geniş ölçekli test verilerinin kolay erişilebilirliği, yenilikçi analiz yöntemlerinin test edilmesine olanak sağlaması ve uluslararası kıyas çalışmalarına imkan vermesi gibi avantajları sebebiyle geniş ölçekli testler konusuna ilgi göstermektedir. Bu ilgi geniş ölçekli testlerle ilgili oldukça fazla sayıda ulusal ve uluslararası yayını da ortaya çıkarmıştır. Bu bağlamlarda değerlendirildiğinde, geniş ölçekli testlerden elde edilen sonuçların ülkelerin eğitim politikalarını etkileme ve yönlendirme gibi çok önemli bir misyonu olduğu açıkça görülmektedir. Makro boyutta politika değişiklikleriyle milyonlarca paydaşı etkileme potansiyeline sahip olan geniş ölçekli test verilerinin bilimsel olarak doğru istatistiksel yöntem ve tekniklerle analiz edilmesi büyük önem arz etmektedir. Çok sayıda değişken içeren ve kendine has özellikleri olan bu verilerin analizi uzmanlık gerektirmektedir. Bu araştırma PISA, PIAAC ve TIMSS verilerinin analizinde dikkat edilmesi gereken hususları ve bu hususlara dikkat edilmediğinde hatalı sonuçlar elde edilebileceğini

örneklendirmekte ve ilgili analiz kodlarını sunmaktadır. Bu sayede geniş ölçekli testlerle ilgili çalışacak araştırmacılara yol gösterilmesi hedeflenmektedir.

### **Geniş Ölçekli Testler**

Çok sayıda uluslararası geniş ölçekli testler mevcut olmakla birlikte en yaygın kullanılanları PISA, PIAAC ve TIMSS'tir. Bu testler bu kısımda ana hatları ile tanıtılmaktadır.

#### *Uluslararası öğrenci değerlendirme programı (PISA)*

PISA, OECD tarafından 2000 yılından itibaren her üç yılda bir 15 yaşındaki öğrencilerin matematik, fen ve okuma alanındaki performanslarını ölçmek amaçlı düzenlenen bir değerlendirme programıdır. PISA, öğrencilerin zorunlu eğitimi tamamlandıktan sonra çağdaş toplumlarda kullanacağı bilgi ve becerilere ne derecede sahip olduğunu ortaya çıkarmayı hedeflemektedir (MEB, 2016a; OECD, 2019). PISA'da üç temel alan bulunmaktadır; okuma becerileri, matematik okuryazarlığı ve fen okuryazarlığı. PISA öğrencinin belirtilen alanlarda öğrendiklerini farklı bağlamlarda ne derece uygulayabildiğini ölçmektedir. PISA, 2009 ve 2018 yıllarında ağırlıklı alan olarak okuma becerilerine odaklanırken, 2003 ve 2012'de matematik okuryazarlığı, 2006 ve 2015 yılında ise fen okuryazarlığına odaklanmıştır. Ayrıca öğrenci, öğretmen, okul müdürü ve velilerden de veri toplanmaktadır. En son uygulanan PISA 2018'e OECD'ye üye olan ve olmayan 76 ülke katılmıştır. Türkiye 2003 yılından itibaren PISA'ya sürekli olarak katılmaktadır.

#### *Uluslararası yetişkin becerilerinin ölçülmesi programı (PIAAC)*

PIAAC, 16-65 yaş aralığındaki yetişkinlerin toplumsal hayata katılım sağlamaları için gerekli temel bilgi işleme becerilerini ortaya koymayı amaçlamaktadır. Bu programın bir çıktısı olan yetişkin becerileri anketi, yetişkin yeterliklerini üç temel bilgi işleme becerisine odaklanarak değerlendirmeyi amaçlar: sözel beceriler, sayısal beceriler ve teknoloji yoğun ortamlarda problem çözme becerileri. Bu becerilerde yetkin olan yetişkinlerin, modern toplumların yaşadığı teknolojik ve yapısal değişikliklerin yarattığı fırsatlardan en iyi şekilde yararlanabilecekleri öngörülmektedir (OECD, 2016). PIAAC, yetişkin becerileri anketi dışında katılımcıların sosyo-demografik özelliklerine ilişkin bilgilerini içeren kapsamlı bir anketi de içermektedir. PIAAC ilk defa 2011-2012'de 24 ülkenin katılımıyla uygulanmaya başlanmış ve 2014-2015'de ikinci turda dokuz ülkenin daha katılımıyla 33 ülkeye ulaşmıştır. Türkiye ikinci turda çalışmaya katılan dokuz ülke arasında yer almıştır. 2016'da yayınlanan *Skills Matter: Further Results from the Survey of Adult Skills (Beceriler Önemlidir: Yetişkin Becerileri Taramasından Sonuçlar)* raporunda açıklanan sonuçlara göre Türkiye belirgin bir şekilde OECD ortalamasının altında yer almaktadır (OECD, 2016; TEDMEM 2016).

#### *Uluslararası matematik ve fen bilimleri eğilimi araştırması (TIMSS)*

TIMSS, uluslararası düzeyde 4 ve 8. sınıfta öğrenim gören öğrencilerin matematik ve fen bilimleri alanlarında kazandıkları bilgi ve becerilerin değerlendirildiği bir tarama çalışmasıdır (MEB, 2016b; Mullis & Martin, 2017). TIMSS, 1995 yılından beri dört yılda bir olmak üzere 1999, 2003, 2007, 2011, 2015 ve 2019 yıllarında uygulanmıştır. İlk uygulamadan bu yana TIMSS tarama araştırmasına katılan ülkelerin sayısı sürekli artmakta ve 2019 yılında TIMSS için katılımcı ülke sayısının 70 olması beklenmektedir (Mullis & Martin, 2017). Türkiye 1999, 2007, 2011, 2015 ve 2019 yıllarında çalışmaya katılmıştır.

Matematik ve fen bilimleri alanlarındaki bilgi ve becerilerinin değerlendirilmesinde, TIMSS temelde geniş bir çerçevede öğretim programında yer alan hedef ve kazanımlara odaklanır. Bu nedenle TIMSS

öğretim programı modeli, ulusal, sosyal ve eğitim bağlamlarında *hedeflenen program*, ev, okul, öğretmen ve sınıf içi bağlamlarda *uygulanan program* ve öğrenci başarısı ve tutumları bağlamında *kazandırılan program* olmak üzere 3 aşamalı bir yapıyı temel almaktadır. Bu bağlamda TIMSS değerlendirme çerçevesi konu alan bilgisine odaklanan konu boyutu ve düşünme süreçlerine odaklanan bilişsel boyuttan oluşur. Ülkelerin matematik ve fen bilimleri alanlarında öğretim programlarına ilişkin sağladığı bu ayrıntılı bilgiler sayesinde TIMSS, ülkelere hem uluslararası karşılaştırma imkanı, hem de ulusal ve yerel çapta değerlendirme yapma fırsatı sunmaktadır (MEB, 2016b).

### **Geniş Ölçekli Test Verilerinin Özellikleri**

Geniş ölçekli test verilerinde dikkate alınması gereken iki temel konu vardır. Bunlar, kullanılan örneklem seçme metoduna bağlı ortaya çıkan örneklem ağırlıkları ve eksik test deseni kullanılmasından dolayı hazırlanan olası değerlerdir (Rutkowski, Gonzalez, Joncas, & von Davier, 2010). Bu kısımda örneklem ağırlıkları ve olası değerler açıklanmaktadır.

#### *Örneklem ağırlıkları*

Geniş ölçekli testlerde evrenin tümünü araştırmaya dahil etmek zamansal ve finansal yetersizlikler gibi sebeplerden ötürü mümkün olmadığından, evrene en iyi genellenebilecek örnekleme seçmek hedeflenmektedir. Örneklem, evrenin temel özelliklerini kestirebildiği ölçüde kullanışlıdır. Örneklem belirleme yöntemlerinin bir sonucu olarak, örneklemin sahip olduğu özelliklerin dağılımının, örneklemin alındığı evrenden farklı olma olasılığı problemini çözümlenen en bilindik yöntem örneklem ağırlıklarının kullanılmasıdır (Rust, 2013). Ulusal düzeyde TÜİK'in yaptığı çalışmalarda ve uluslararası geniş ölçekli testlerde (PISA, TIMSS, PIRLS ve PIAAC gibi) örneklem ağırlıkları kullanılmaktadır. Örneğin, PISA 2015 teknik raporunda, örneklemdaki her öğrencinin tüm PISA evreninde yer alan gerçek öğrenci sayısını en doğru şekilde temsil ettiğinden emin olmanın gerekliliğinden bahsedilerek, PISA verilerinin analizinde örneklem ağırlıklarının kullanılması gerektiği vurgulanmaktadır (OECD, 2017).

PISA ve TIMSS geniş ölçekli testlerin örneklem seçiminde çok aşamalı (multistage) desen kullanılmaktadır. Çok aşamalı desenlerin kullanılması, analizler sonucunda elde edilen kestirimlerin kesinliğine önemli ölçüde etki edebilmektedir (Rust, 2013). İlk aşamada okullar büyüklükleri ile doğru orantılı bir olasılıkla seçilir; ikinci aşamada da belirlenen okullardan sınıflar ve/veya öğrenciler rastgele seçilir (LaRoche & Foy, 2016; OECD, 2017). Okulun büyüklüğü çalışmaya katılabilecek öğrenci sayısı ile belirlenir. Örneğin PISA için okuldaki 15 yaşındaki öğrencilerin sayısı, TIMSS için 4. sınıfa veya 8. sınıfa kayıtlı öğrenci sayısı okul büyüklüğünü hesaplamada kullanılır. PIAAC için ise 16-65 yaş arasında ülkede ikamet eden, akli dengesi yerinde, bakıma muhtaç olmayan, tutuklu olmayan ve askeriyede görevli olmayan tüm yetişkinler dikkate alınır.

Örneklem belirlenirken seçim sürecinin yanlı olmaması ve her bireyin örnekleme seçilme şansının eşit olması için mümkünse rastgele seçim (random sampling) yöntemi uygulanır. Rastgele olmayan örneklem seçimleri isteyerek ya da istemeyerek de olsa yanlılıklara sebep olabilmektedir. Rastgele seçimde de her zaman evrende her bireyin örnekleme seçilme şansı eşit olmayabilir. Bu durumda yanlılıktan kaçınmak her bireyin evrendeki temsiliyetini garantilemek için örneklem ağırlıkları kullanılır. Örneklem içindeki her bireyin seçilme olasılığına göre bir örneklem birimi belirlenir. Örneklem ağırlıkları ise bir birimin seçilme olasılığının tersi olarak tanımlanmaktadır. Bir başka deyişle, eğer bir özelliğe sahip bir grubun örnekleme seçilme şansı çok düşükse o grubu temsil eden bireyin örneklem birimi, seçilme şansı çok yüksek gruptan gelen bir bireyin örneklem biriminden daha

büyük ve önemli olacaktır (OECD, 2017). Yapılan analizlerde grupların puan ortalamaları için örneklem ağırlıkları hesaba katıldığında evren temsiliyeti korunmuş olur ve doğru kestirimler yapılır. Örneklemeden elde edilen veriyi analiz ederken, örneklem ağırlıkları kullanıldığı zaman her bir öğrencinin istatistiksel kestirimlere olan katkısı, evrende temsil edilen öğrencilerin sayısı ile orantılı olmaktadır (Gonzales, 2012). Örneğin 300 kişiden her bir bireyin seçilme şansının eşit olduğu durumda rastgele seçilecek 30 kişinin her birinin seçilme olasılığı  $1/10$  ve her bir bireyin ağırlığı da 10 olmuş olur. Bu örnekte her bireyin seçilme olasılığı eşit olduğundan her birinin örneklem ağırlığı da eşittir ve 30 kişinin ağırlıklarını topladığımızda evren sayısı 300'e ulaşmış oluruz. Böyle bir durumda ağırlıklandırılmış ortalama ile ağırlıklandırılmamış ortalama birbirine eşit olmaktadır. Örneğin 45 kişilik bir sınıfta 15 kız 30 erkek öğrencinin olduğu bir evrenden 6 öğrencinin oluşturduğu bir örneklemin rastgele seçildiğini düşünelim. Bu seçim sonucu 3 kız 3 erkek öğrenci seçilmiş olsun. Evrende erkekler kızlardan daha çok temsil edilirken örneklemede eşit temsil edilmektedirler. 15 kız öğrenci içinden 3 kız öğrencinin her birinin seçilme olasılığı  $3/15 = 0,2$  ve 30 erkek öğrenci içinde 3 erkek öğrencinin her birinin seçilme olasılığı da  $3/30 = 0,1$  olacaktır. Bu durumda örneklemedeki her bir kız öğrencinin ağırlığı 5 iken her bir erkek öğrencinin ağırlığı da 10 olur. Bu öğrencilerin 10 puan üzerinden hesaplanan bir sınavdan aldıkları notlar kız öğrencilerin 8, 7, 7 olsun ve erkek öğrencilerin aldıkları notlar ise 5, 5, 4 olsun. Bu durumda örneklemin ağırlıklandırılmamış ortalaması  $[(8 + 7 + 7) + (5 + 5 + 4)]/6 = 6$  iken ağırlıklandırılmış ortalama her birinin puanının ağırlıkları ile çarpımının toplam ağırlığa bölünmesinden elde edilen sonuç yani  $[(8 \times 5 + 7 \times 5 + 7 \times 5) + (5 \times 10 + 5 \times 10 + 4 \times 10)]/45 = 5,56$  olur. Ağırlıklandırılmış ortalama ağırlıklandırılmamış ortalamadan %7 daha düşük olmuştur. En basit şekli ile bu örnekte görüldüğü gibi ağırlıkları hesaba katmadan yapılan analizler sonucu evren ile ilgili yapılacak olan kestirimler yanıltıcı olacaktır.

Çok aşamalı örneklem belirleme deseninde, önce okulların belirlendiği daha sonra okul içinden öğrencilerin seçildiği bir uygulamada, okul ağırlığı, okul içi ağırlık ve öğrenci ağırlığı ayrı ayrı belirlenmektedir. Örneğin,  $j$  okulunun seçilme olasılığı  $p_j$  ve  $j$  okulundaki  $i$  öğrencisinin,  $j$  okulu seçildiği koşulda seçilme olasılığı  $p_{ij}$  olsun. Öğrenci düzeyinde okul içi ağırlık  $w_{ij} = 1/p_{ij}$  olurken, okul düzeyindeki ağırlık  $w_j = 1/p_j$  olur. Her birinde 40 öğrencinin olduğu 10 farklı okuldaki 400 öğrencilik bir evrende, önce 4 okul rastgele seçilsin. Sonra her okul içinden de 10'ar öğrenci seçilsin. Toplamda örnekleme oluşturan öğrenci sayısı 40 olur. Bu durumda her bir okulun seçilme olasılığı (10 okuldan 4 okul seçiliyor);  $p_j = 4/10 = 0,4$  olduğundan okul ağırlığı  $w_j = 2,5$  olmaktadır. Seçilen 4 okul içinden her bir öğrencinin seçilme olasılığı (her okuldaki 40 öğrenci içinden 10 öğrenci seçiliyor)  $p_{ij} = 10/40 = 0,25$  ve okul içi ağırlıklandırma  $w_{ij} = 4$  olur. Son olarak öğrenci bazında her bir öğrencinin; önce okulun, sonra okul içinden öğrencilerin seçildiği durumda son seçilme olasılığı;  $p_{*ij} = p_j \times p_{ij} = 0,4 \times 0,25 = 0,10$  ve öğrenci ağırlığı  $w_{*ij} = 10$  olur.

PISA, PIAAC ve TIMSS gibi geniş ölçekli testlerin verileri örneklem ağırlıkları kullanılarak elde edildiği için, yapılacak tüm analizlerde bu ağırlıkları dikkate alan yöntem ve yazılımlar kullanılması zorunludur. Öğrenci ağırlıkları PISA'da W\_FSTUWT (Final trimmed nonresponse adjusted student weight), PIAAC'da SPFWT0 (Final full sample weight) ve TIMSS'de TOTWGT (Total student weight) değişken isimleri ile veri setinde yer almaktadır. Çok düzeyli analizlerde (öğrenci, sınıf, okul gibi) örneklem ağırlıkları ilgili düzeylere göre ayrıştırılarak kullanılmalıdır (Rutkowski vd., 2010). Örneklem ağırlıklarını dikkate almadan elde edilen sonuçların hatalı olacağına farkında olunması gerekmektedir (LaRoche & Foy, 2016; OECD, 2017, Rutkowski vd., 2010). Rutkowski vd. (2010) TIMSS 2007 Bulgaristan ortalama matematik puanının örneklem ağırlıkları kullanıldığında 463,63, örneklem ağırlıkları kullanılmadığında 481,38 olduğunu hesaplamışlardır.

### *Olası değerler*

PISA, PIAAC ve TIMSS gibi geniş ölçekli değerlendirmelerin amacı bireylerin aldıkları puandan ziyade evrenin veya içerisindeki alt grupların performansını tanımlamaktır (Monseur & Adams, 2009; Von Davier, Gonzalez, & Mislavy, 2009). Bireyler için tutarlı ve geçerli puanlar hesaplamak geniş ölçekli testlerin amacı değildir. Bu sebeple kestirimler yapılırken evren seviyesinde hataların azaltılması amaçlanmaktadır (OECD, 2017). Ayrıca bireyler üzerindeki test yükünü hafifletmek amacı ile de eksik test deseni (rotated booklet design) kullanılmaktadır (Rutkowski vd., 2010). Bir öğrenci soruların sadece kısıtlı bir kısmını yanıtlamaktadır. Grup olarak ise öğrenciler tüm soruları yanıtlamış olurlar. Bu sebeplerle geniş ölçekli testlerde test sonuçları olası değerler (plausible values-PVs) olarak raporlanmaktadır.

Olası değerler yöntemi öğrenci başarısını kayıp değer olarak kabul eder (Rutkowski vd., 2010). Öğrenci başarı dağılımları Rubin'in (1987) çoklu atama (multiple imputation) yöntemi ile kestirilir. Bu dağılımların içinden rastgele seçimler yapılır ve atanan çoklu verilere olası değerler denir (Rutkowski vd., 2010). Olası değerler gözlemlenemeyen gizil değişkenler için emsal değerlerdir (Wu, 2005). Her öğrencinin gözlemlenemeyen gizil başarı değişkeni vardır ve bu değişkene ait çoklu veriler atanmaktadır. (Laukaityte & Wiberg, 2017; Wu, 2005). OECD (2017) olası değerleri, puan dağılımlarından bireylere atamak için rastgele seçilen sayılar olarak tanımlamaktadır. Bu dağılıma ise marjinal sonsal dağılım (marginal posterior distribution) denmektedir. Seçkisiz hata varyans bileşenleri içeren olası değerler test puanları olarak ele alınmadan, evrenin performansını tanımlamak amaçlı kullanılmaktadır (OECD, 2017). Kısacası ölçüm hatalarını azaltmak için her kişiye birden fazla değer atanmaktadır (Laukaityte & Wiberg, 2017). Eğer ölçüm hatası küçükse, kişi için atanan çoklu değerler birbirine yakın; ölçüm hatası büyük ise de atanan değerler birbirine uzak olmaktadır (Wu, 2005). Atanan olası değerler sayesinde geniş ölçekli testlerdeki çıkarımlar daha geçerli hale gelmekte ve sonuçlarının pratiğe katkısı daha verimli olmaktadır (Laukaityte & Wiberg, 2017).

PISA ve TIMSS gibi bir çok geniş ölçekli testlerin verilerinde 5 tane olası değer kullanılmaktadır (Laukaityte & Wiberg, 2017, OECD, 2017). PISA, 2015 uygulamasından beri 10 tane olası değer raporlamaya başlamıştır. PIAAC içerisinde de 10 tane olası değer raporlanmaktadır. National Assessment of Educational Assessment (NAEP) veritabanı ise 20 tane olası değer kullanılmaktadır. Laukaityte ve Wiberg (2017)'in yaptıkları simülasyon çalışmaları birden fazla olası değer kullanılmasının yapılan tahminin ve ölçüm hatasının doğruluğunu arttırdığını göstermiştir.

PISA, PIAAC ve TIMSS gibi geniş ölçekli testlerin verilerinde sonuçlar olası değerler olarak raporlandığı için yapılacak tüm analizlerde bu olası değerleri dikkate alan yöntem ve yazılımların kullanılması zorunludur. Olası değerleri dikkate almadan elde edilen sonuçların hatalı olacağına farkında olunması gerekmektedir (LaRoche & Foy, 2015; OECD, 2017, Rutkowski vd., 2010).

### ***Geniş Ölçekli Testlerin Analizinde Yapılan Sistemik Hatalar***

Rutkowski vd. (2010) geniş ölçekli testler analiz edilirken yapılan iki önemli hatayı raporlamıştır. İlk hata, herhangi bir olası değeri olarak analizleri yapmaktır. İkinci hata ise olası değerlerin ortalamasını olarak analizleri yapmaktır. Örneğin TIMSS için ilk olası değeri (PV1) veya herhangi bir olası değeri olarak analizleri yapmak ilk hataya örnektir. Rutkowski vd. (2010) ortalamaları olarak analize katmanın çok daha büyük problemler yaratabileceğini belirtmiştir. Ortalama puanlar kesinlikle kullanılmamalıdır diye de uyarmıştır. Her iki durumda da standart hatalar farklı kestirilecek ve anlamlılık değeri etkilenecektir. Ayrıca, olası değerlerin bir örtük değişkenin farklı boyutları gibi olarak yapısal eşitlik modellemesinde kullanmak da hatalıdır. Bunlara ek olarak örneklem ağırlıklarının kullanılmaması da bir hata kaynağıdır. Türkiye'de yapılan çalışmalara bakıldığında doğru teknikleri kullanan çalışmalar olduğu gibi olası değerlerden birini alan veya ortalamasını alan

çalışmalar da mevcuttur. Bu makalede aktarılan bilgiler ve verilen örnekler baz alınarak doğru yöntemlerin kullanılmasının yaygınlaşması amaçlanmaktadır.

### ***Araştırmanın Amacı***

Bu çalışmanın amacı uluslararası geniş ölçekli testlerin doğru yöntemlerle analiz edilmesi gerektiğini nedenleri ile açıklayarak temel analizleri örneklendirmektir. Bu amaçla geniş ölçekli test verileri kullanılarak farklı grupların karşılaştırılmasını örneklendiren t-testi, seçilen bir bağımlı değişkeni yordamayı örneklendiren regresyon analizi ve çok düzeyli veri yapılarının analizini örneklendiren çok düzeyli regresyon (multilevel regression) analizini içeren örnek araştırma soruları seçilmiştir. Bu analizlerin yapılabilmesi için gerekli olan kodlar EK A, B, C ve D’de verilmektedir. Bu çalışmaya yön veren araştırma soruları altta verilmektedir.

- 1) Geniş ölçekli test verilerinde örneklem ağırlıkları ve olası değerler dikkate alınmadığında t-testi sonuçları nasıl etkilenmektedir?
- 2) Geniş ölçekli test verilerinde örneklem ağırlıkları ve olası değerler dikkate alınmadığında çoklu regresyon sonuçları nasıl etkilenmektedir?
- 3) Geniş ölçekli test verilerinde örneklem ağırlıkları ve olası değerler dikkate alınarak çok düzeyli regresyon analizleri nasıl gerçekleştirilir?

Birinci araştırma sorusunu yanıtlamak için “TIMSS 2015 uygulamasında Türkiye’deki kız ve erkek öğrencilerin matematik puan ortalamaları arasında istatistiksel olarak anlamlı bir fark var mıdır?” ve “PIAAC 2015 uygulamasında Türkiye’deki yetişkinlerden bir önceki ay iş arayanlarla iş aramayanların okuma puan ortalamaları arasında istatistiksel olarak anlamlı bir fark var mıdır?” alt araştırma soruları incelenmiştir. İkinci araştırma sorusunu yanıtlamak için “PISA 2015 uygulamasında fen dersindeki disiplin iklimi, epistemolojik inançlar, sosyoekonomik statü, araştırma temelli fen öğretimi, araçsal motivasyon, feni sevme, fende özyeterlik, öğretmen merkezli fen öğretimi, öğrenci isteğine bağlı öğretmen desteği değişkenleri birlikte Türkiye’deki öğrencilerin fen performans puanlarını anlamlı bir şekilde yordamakta mıdır?” alt araştırma sorusu incelenmiştir. Üçüncü araştırma sorusunu yanıtlamak için “TIMSS 2011 uygulamasında öğrenci düzeyi değişkenleri ailenin ödev zaman ayrılmasını sağlaması, ailenin ödevin yapıldığını kontrol etmesi ve öğrencinin ödevi ayırdığı süre; öğretmen düzeyi değişkenleri öğretmenin ödevi düzeltmesi ve geri dönüt vermesi, öğrencilere kontrol ettirmesi, ödevi tartışması, tamamlanmasını izlemesi, not için kullanması değişkenleri akıl yürütme puanlarını anlamlı bir şekilde yordamakta mıdır?” alt araştırma sorusu incelenmiştir.

## **YÖNTEM**

### ***Örneklem***

Bu çalışmada çeşitliliği sağlamak adına PISA, PIAAC ve TIMSS veri setleri kullanılmıştır. Bu veri setlerinde yer alan örneklem bu kısımda açıklanmaktadır. PISA 2015 uygulamasına Türkiye’deki 187 okulda yer alan 5895 öğrenci katılmıştır. Öğrencilerin çoğunluğu 10. sınıf olan öğrencilerdir (MEB, 2016a). PIAAC 2015 uygulamasına Türkiye’den 16 ile 65 yaş aralığına yer alan 5227 yetişkin katılmıştır (OECD, 2016). TIMSS 2011 uygulamasına 8. sınıflar seviyesinde 239 okulda yer alan 6928 öğrenci katılmıştır (MEB, 2014). TIMSS 2015 uygulamasına ise 8. sınıflar seviyesinde 238 okulda yer alan 6079 öğrenci katılmıştır (MEB, 2016b).

### Ölçme Aracı

PISA, PIAAC ve TIMSS hem başarı testleri hem de anket boyutları olan uygulamalardır. Bu sayede araştırmacılara çok çeşitli özellikleri ölçen veriler sağlamaktadırlar. Bu kısımda çalışmada kullanılan değişkenler tanımlanmaktadır.

Birinci araştırma sorusunda TIMSS 2015 matematik puanları 5 olası değer olarak raporlanmaktadır (BSMMAT01-BSMMAT05). Puanlar madde tepki kuramı kullanılarak kestirilmiştir. Araştırma sorusunun diğer değişkeni olan cinsiyet bilgisi anket verisinden alınmıştır (BSBG01).

Tablo 1. Çalışmada Kullanılan TIMSS 2015 Değişkenleri

Değişken	Verideki Tanımlama İsmi
Cinsiyet	BSBG01
Başarı Puanı	BSMMAT01, BSMMAT02, BSMMAT03, BSMMAT04, BSMMAT05

PIAAC 2015 verisinin kullanıldığı araştırma sorusunda yetişkinlerin okuma puanları ve bir önceki ay iş arama durumları değişkenleri kullanılmaktadır. Okuma puanları 10 farklı olası değerle raporlanmaktadır (PVLIT1- PVLIT10). Okuma testine verilen yanıtlara göre okuma puanları madde tepki kuramı kullanılarak kestirilmiştir. İş arama durumu ise uygulanan ankete verilen yanıtlardan elde edilmiştir (C\_Q02b).

Tablo 2. Çalışmada Kullanılan PIAAC 2015 Değişkenleri

Değişken	Verideki Tanımlama İsmi
İş Arama Durumu	C_Q02b
Başarı Puanı	PVLIT1, PVLIT2, PVLIT3, PVLIT4, PVLIT5, PVLIT6, PVLIT7, PVLIT8, PVLIT9, PVLIT10

PISA 2015 ile ilgili olan araştırma sorusunda disiplin iklimi, epistemolojik inançlar, sosyoekonomik statü, araştırma temelli fen öğretimi, araçsal motivasyon, feni sevme, fende özyeterlik, öğretmen merkezli fen öğretimi, öğrenci isteğine bağlı öğretmen desteği değişkenleri bağımsız değişken olarak kullanılmıştır (DISCLISCI, EPIST, ESCS, IBTEACH, INSTSCIE, JOYSCIE, SCIEEFF, TDTEACH, TEACHSUP). Bu değişkenler anket sorularına öğrencilerin verdiği yanıtlar kullanılarak elde edilen standartlaştırılmış ölçek puanlarıdır. İlgili boyutları temsil eden anket soruları madde tepki kuramı kullanılarak ölçek puanına dönüştürülmüşlerdir. PISA 2015'te fen performansı 10 farklı olası değerle raporlanmaktadır (PV1SCIE-PV10SCIE). Fen testine verilen yanıtlara göre fen puanları madde tepki kuramı kullanılarak kestirilmiştir.



Tablo 3. Çalışmada Kullanılan PISA 2015 Değişkenleri

Değişken	Verideki Tanımlama İsmi
Disiplin İklimi	DISCLISCI
Epistemolojik İnançlar (İnanç)	EPIST
Sosyoekonomik Statü (SES)	ESCS
Araştırma Temelli Fen Öğretimi (Araştırma T. Fen Ö.)	IBTEACH
Araçsal Motivasyon (Motivasyon)	INSTSCIE
Feni Sevme	JOYSCIE
Fende Özyeterlik (Öz yeterlik)	SCIEEFF
Öğretmen Merkezli Fen Öğretimi (Öğret. M. Fen. Ö.)	TDTEACH
Öğrenci İsteğine Bağlı Öğretmen Desteği (Destek)	TEACHSUP
Başarı Puanı	PV1SCIE, PV2SCIE, PV3SCIE, PV4SCIE, PV5SCIE, PV6SCIE, PV7SCIE, PV8SCIE, PV9SCIE, PV10SCIE

TIMSS 2011 ile ilgili olan sonuncu araştırma sorusunda ise hiyerarşik yapıya sahip olan değişkenler kullanılmıştır. Öğrenci düzeyinde öğrencilerin anketlere verdikleri yanıtlara göre ailenin ödev zaman ayrılmasını sağlaması, ailenin ödevin yapıldığını kontrol etmesi ve öğrencinin ödevi ayırdığı süre değişkenleri elde edilmiştir (BSBG11C, BSBG11D, BSBM20B). Öğretmen düzeyinde ise öğretmenlerin anketlere verdikleri yanıtlara göre öğretmenin ödevi düzeltmesi ve geri dönüt vermesi, öğrencilere kontrol ettirmesi, ödevi tartışması, tamamlanmasını izlemesi, not için kullanması değişkenleri kullanılmıştır (BTBM25CA, BTBM25CB, BTBM25CC, BTBM25CD, BTBM25CE). Bağımlı değişken olan akıl yürütme puanları ise ilgili sorulara verilen yanıtlara göre madde tepki kuramı kullanılarak 5 olası değer olarak kestirilmektedir (BSMREA01-BSMREA05).

Tablo 4. Çalışmada Kullanılan TIMSS 2011 Değişkenleri

Değişken	Verideki Tanımlama İsmi
Ailenin Ödev Zaman Ayrılmasının Sağlanması	BSBG11C
Ailenin Ödevin Yapıldığını Kontrol Etmesi	BSBG11D
Öğrencinin Ödevi Ayırdığı Süre	BSBM20B
Öğretmenin Ödevi Düzeltmesi ve Geri Dönüt Vermesi	BTBM25CA
Öğretmenin Ödevi Öğrencilere Kontrol Ettirmesi	BTBM25CB
Öğretmenin Ödevi Tartışması	BTBM25CC
Öğretmenin Ödevin Tamamlanmasını İzlemesi	BTBM25CD
Öğretmenin Ödevi Not İçin Kullanması	BTBM25CE
Başarı Puanı	BSMREA01, BSMREA02, BSMREA03, BSMREA04, BSMREA05

### Veri Analizi

Bu kısımda gerçekleştirilen analiz yöntemleri ve bu analizlerde dikkat edilmesi gereken hususlar aktarılmaktadır. Birinci araştırma sorusu gruplar arası ortalama karşılaştırma analizidir. İki bağımsız grup olduğu için bağımsız örneklem t-testi kullanılan bu araştırma sorusunda, normallik ve eşvaryanslılık varsayımlarının sağlandığı varsayılarak analizler gerçekleştirilmiştir. Giriş kısmında da anlatıldığı gibi analizleri yaparken örneklem ağırlıklarına ve olası değerleri dikkate alan programlar kullanılmalıdır. IEA'nın (International Association for Evaluation) IDB Analyzer programı geniş ölçekli test verilerinde örneklem ağırlıklarının ve olası değerleri dikkate alarak t-testi yapabilen bir programdır (IEA, 2019). IDB Analyzer programı SPSS dosyasını okuyabilen bir ara programdır. IDB Analyzer kullanılırken analizde kullanılan değişkenler, olası değerler dahil, ilk aşamada seçilir. Ardından kullanılan verisetine uygun örneklem ağırlıklarının seçilmesi gerekmektedir. Bu işlemlerin sonunda IDB Analyzer programı SPSS kodu üretmekte ve üretilen kodun çalıştırılması ile sonuçlar

elde edilmektedir. IDB Analyzer sonuç dosyasında anlamlılık değerini (p değeri) vermemekte ancak t istatistiklerini vermektedir. İlgili serbestlik derecesi baz alınarak, elde edilen farkın veya ilişkilerin istatistiksel olarak anlamlı olup olmadığına karar verilmektedir. Bu değerler oluşan dosyalar arasında “\*\_sig.sav” dosyasında bulunabilir. TIMSS ile ilgili analiz öğrenci seviyesinde olduğu için “Total Student Weight” (TOTWGT), PIAAC ile ilgili analiz yetişkin seviyesinde olduğu için “Final Full Sample Weight” (SPFWT0) kullanılmıştır.

İkinci araştırma sorusu birden fazla bağımsız değişkeni kullanarak bir bağımlı değişkeni yordamak ile ilgili olduğu için çoklu regresyon analizi yapılmıştır. Normallik, doğrusallık, çoklu doğrusal bağlantı problemi, artıkların eşvaryanslılığı ve artıkların bağımsızlığı varsayımlarının sağlandığı kabul edilmiştir. Bu analizde de benzer şekilde IDB Analyzer programı örneklem ağırlıklarını ve olası değerleri dikkate alabildiği için kullanılmıştır. PISA 2015 verisinde öğrenci düzeyi örneklem ağırlığı olarak FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT (W\_FSTUWT) kullanılmıştır.

Sonuncu araştırma sorusunda ise çok düzeyli bir yapı olduğu için çok düzeyli regresyon analizi yapılmıştır. Bu analizde doğrusallık, artıkların bağımsızlığı, artıkların eşvaryanslılığı ve rastgele kesenler varsayımlarının sağlandığı varsayılmıştır. Mplus programı da örneklem ağırlıklarını ve olası değerleri dikkate alabilmektedir (Muthen & Muthen, 2015). Örneklem ağırlıkları yazılan kodda tanımlanmalıdır. Rutkowski vd. (2010) önerdiği gibi çok düzeyli olan bu araştırma sorusunda kullanılması gereken ağırlıklar ayrıca araştırmacılar tarafından hesaplanmıştır. TIMSS örneklem seçiminde önce okul, sonra sınıf ve sınıftaki tüm öğrenciler seçildiği için birinci düzey örneklem ağırlığı  $WGTADJ2 * WGTFAC2 * WGTADJ3 * WGTFAC3$  çarpımı sonucu hesaplanmıştır (CLASS WEIGHT ADJUSTMENT \* CLASS WEIGHT FACTOR \* STUDENT WEIGHT ADJUSTMENT \* STUDENT WEIGHT FACTOR). İkinci düzey için örneklem ağırlığı  $WGTADJ1 * WGTFAC1$  (SCHOOL WEIGHT ADJUSTMENT \* SCHOOL WEIGHT FACTOR) çarpımı ile elde edilmiştir. Burada elde edilen birinci düzey ağırlık ve ikinci düzey ağırlığın çarpımı ise TIMSS tarafından raporlanan toplam öğrenci ağırlığını vermektedir (TOTAL STUDENT WEIGHT). MPLUS programında olası değerlerin kullanılması için ise her olası değer için ayrı ayrı bir kez tüm diğer değişkenlerle birlikte yer aldığı veri dosyalarının oluşturulması gerekmektedir. Örneğin 5 olası değer varsa olası değer için yer aldığı sütun dışındaki tüm sütunların aynı olduğu 5 veri dosyası oluşturulmalıdır. Bu 5 veri dosyasının adı farklı bir veri dosyasında alt alta sıralanmalıdır. Mplus kodunda ise bu veri dosyası isimlerini içeren liste dosya çağrılmalıdır (FILE = dataimputedlist.dat;). Ayrıca, bu yapıda veri dosyası kullanıldığı da kodlarda belirtilmelidir (TYPE = IMPUTATION;). Ardından değişkenler arası ilişkiler tanımlanmalıdır.

## BULGULAR

Bulgular kısmında grupların ortalama puanlarını karşılaştıran, bir bağımlı değişkeni tek düzeyde yordayan ve bir bağımlı değişkeni çok düzeyli veri setleri ile yordayan çalışmalara yer verilmektedir.

### *Grupların Ortalama Puanlarını Karşılaştırılmayı Amaçlayan Çalışmalar*

Bu kısımda iki alt araştırma sorusu yer almaktadır. İlk olarak “TIMSS 2015 uygulamasında Türkiye’deki kız ve erkek öğrencilerin matematik puan ortalamaları arasında anlamlı bir fark var mıdır?” araştırma sorusu yanıtlanmaktadır. Kategorik iki grubun sürekli puan ortalamaları karşılaştırıldığı için t-testi tercih edilmiştir. Örneklem ağırlıkları ve olası değerleri dikkate alarak ve almayarak elde edilen sonuçlar Tablo 5’te verilmektedir.

Elde edilen sonuçlara göre kızlar ve erkeklerin matematik ortalamaları arasında istatistiksel olarak anlamlı bir fark bulunmamaktadır ( $t=1,79$ ,  $p>,05$ ). Bu sonuç TIMSS 2015 Ulusal Matematik ve Fen Ön Raporu tarafından da doğrulanmaktadır (MEB, 2016b).

Tablo 5'te ayrıca olası değerlerin tek tek kullanıldığı veya olası değerlerin ortalaması alınarak kullanıldığı durumlarda elde edilen sonuçlar verilmektedir. Örneklem ağırlıkları dikkate alınmadan olası değerler tek tek kullanıldığında veya ortalaması alınarak kullanıldığında kızlar ve erkekler arasında istatistiksel olarak anlamlı fark bulunmaktadır. Özetle, uygun analiz yöntemi kullanmadan gruplar arası fark analizi yapıldığında hatalı sonuçlar bulmak mümkündür. Bu araştırma sorusu bu durumu net bir şekilde örneklendirmektedir.

Tablo 5. Kızlar ve Erkeklerin Matematik Puan Ortalamaları Karşılaştırması

Yöntem	Kız (Ort) (SH)	Erkek (Ort) (SH)	Ortalama farkı (SH)	<i>t</i>
IDB Analyzer PV1-PV5	461.14 (4.80)	454.73 (5.31)	6.40 (3.57)	1.79
SPSS PV1	459.23 (1.90)	452.77 (1.86)	6.46 (2.66)	2.43*
SPSS PV2	460.50 (1.91)	452.87 (1.87)	7.63 (2.67)	2.85**
SPSS PV3	460.26 (1.91)	451.33 (1.91)	8.93 (2.70)	3.31**
SPSS PV4	458.04 (1.97)	449.01 (1.94)	9.03 (2.77)	3.26**
SPSS PV5	459.37 (1.94)	453.84 (1.90)	5.53 (2.72)	2.04*
SPSS PVort	459.48 (1.87)	451.97 (1.83)	7.51 (2.62)	2.87**

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ . Ort: Ortalama. SH: Standart Hata.

Diğer alt araştırma sorusunda PIAAC uygulamasına odaklanılmıştır. “PIAAC 2015 uygulamasında Türkiye’deki yetişkinlerden bir önceki ay iş arayanlarla iş aramayanların okuma puan ortalamaları arasında istatistiksel olarak anlamlı bir fark var mıdır?” araştırma sorusunda benzer şekilde ortalamalar karşılaştırılmaktadır.

Tablo 6’da ilk olarak örneklem ağırlıklarını ve olası değerleri dikkate alan IDB Analyzer ile elde edilen sonuçlar raporlanmaktadır. Elde edilen sonuçlara göre iş arayanlarla iş aramayanların okuma becerisi ortalamaları arasında istatistiksel olarak anlamlı bir fark bulunmamaktadır ( $t=1,16$ ,  $p>,05$ ).

Örneklem ağırlıkları dikkate alınmadan olası değerler teker teker kullanıldığında veya ortalaması alınarak kullanıldığında 11 farklı değerde (10 olası değer ve olası değerlerin ortalamaları) birbiri ile çelişen sonuçlar elde edilmektedir. Üç değerde istatistiksel olarak anlamlı fark bulunurken sekiz değerde fark yok sonucu ortaya çıkmıştır. Özetle, uygun analiz yöntemi kullanmadan grupların ortalamaları karşılaştırıldığında hatalı sonuçlar bulmak mümkündür.

Tablo 6. İş Arayanlarla İş Aramayanların Okuma Puan Ortalamaları Karşılaştırması

Yöntem	İş Arayanlar (SH)	İş Aramayanlar (SH)	Ortalama farkı (SH)	<i>t</i>
IDB Analyzer PV1-PV10	226.11 (4.16)	221.05 (1.45)	5.06 (4.36)	1.16
SPSS PV1	229.06 (2.51)	223.90 (.83)	5.16 (2.73)	1.89
SPSS PV2	229.40 (2.59)	223.23 (.83)	6.17 (2.75)	2.25*
SPSS PV3	227.01 (2.57)	224.33 (.83)	2.67 (2.73)	.98
SPSS PV4	226.87 (2.45)	224.12 (.84)	2.76 (2.74)	1.01
SPSS PV5	226.52 (2.55)	222.94 (.83)	3.58 (2.71)	1.32
SPSS PV6	231.42 (2.58)	224.81 (.84)	6.61 (2.75)	2.40*
SPSS PV7	226.62 (2.55)	223.93 (.82)	2.70 (2.71)	1.00
SPSS PV8	226.73 (2.56)	223.70 (.83)	3.03 (2.72)	1.11
SPSS PV9	227.88 (2.51)	222.49 (.84)	5.39 (2.76)	1.95
SPSS PV10	231.14 (2.63)	222.82 (.84)	8.32 (2.75)	3.02**
SPSS PVort	228.27 (2.34)	223.63 (.76)	4.64 (2.51)	1.85

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ . Ort: Ortalama. SH: Standart Hata.

İki analiz yöntemi arasında ortaya çıkan bu farkın genel olarak standart hatalardan kaynaklandığı görülmektedir. Doğru yöntemlerle hesaplanan standart hatalar daha yüksek iken, örneklem ağırlıklarını ve olası değerleri dikkate almadan yapılan hesaplamalarda standart hata daha düşüktür. Bu fark hatalı *t* değerleri elde edilmesine ve verilen kararın hatalı olmasına neden olmaktadır.

### **Bir Bağımlı Değişkeni Tek Düzeyde Yordamayı Amaçlayan Çalışmalar**

Bu kısımda “PISA 2015 uygulamasında fen dersindeki disiplin iklimi, epistemolojik inançlar, sosyoekonomik statü, araştırma temelli fen öğretimi, araçsal motivasyon, feni sevme, fende özyeterlik, öğretmen merkezli fen öğretimi, öğrenci isteğine bağlı öğretmen desteği değişkenleri birlikte Türkiye’deki öğrencilerin fen performans puanlarını anlamlı bir şekilde yordamakta mıdır?” alt araştırma sorusu yanıtlanmaktadır. Bağımlı bir değişkeni yordayan bağımsız değişkenlerin tespiti amaçlandığı için çoklu regresyon analizi kullanılmıştır. Örneklem ağırlıklandırma ve olası değerleri dikkate alarak ve almayarak elde edilecek sonuçlar Tablo 7’de verilmektedir.

Örneklem ağırlığı ve olası değerler dikkate alınarak yapılan analiz sonuçlarına göre araçsal motivasyon ve öğrenci isteğine bağlı öğretmen desteği değişkenlerinin fen performansını yordamada rolleri yoktur. Fen dersindeki disiplin iklimi, epistemolojik inançlar, sosyoekonomik statü, araştırma temelli fen öğretimi, feni sevme, fende özyeterlik ve öğretmen merkezli fen öğretimi değişkenleri fen performansını yordamaktadır.

Örneklem ağırlığı ve olası değerler dikkate alınmadan yapılan analiz sonuçlarına göre 11 verisetinin 8’i hatalı sonuçlar vermektedir. İstatistiksel olarak anlamlı çıkmaması gereken ilişkiler hatalı olarak istatistiksel olarak anlamlı bulunmaktadır. Bu durum yine standart hatanın yanlış kestirilmesinden yani olması gerekenden küçük hesaplanmasından kaynaklanmaktadır. Yaygın olarak yapılan hatalardan

olan PV1'in veya PV ortalamasının kullanılması durumlarının ikisi de hatalı sonuç vermiştir. Bu örnekte  $R^2$  değerlerinin çok değişmediği ama PVortalama için elde edilen değer genelden daha yüksek olduğu görülmektedir. Bu örnek de göstermektedir ki örneklem ağırlığı ve olası değerler regresyon analizinde de dikkate alınmalıdır.

Tablo 7. Fen Başarısını Yordayan Faktörler

Yöntem	Disiplin	İnanç	SES	Araştırma T. Fen Ö.	Motivasyon	Feni Sevme	Öz yeterlik	Öğret. M. Fen Ö.	Destek	$R^2$
IDB Analyzer PV1-PV10	.09***	.19***	.27***	-.19***	.03	.09***	.08***	.04*	.03	.20
SPSS PV1	.08***	.19***	.26***	-.18***	.03*	.09***	.08***	.05***	.03*	.19
SPSS PV2	.07***	.20***	.26***	-.18***	.03*	.11***	.08***	.04**	.02	.20
SPSS PV3	.09***	.20***	.27***	-.19***	.03*	.09***	.08***	.05***	.02	.20
SPSS PV4	.09***	.19***	.26***	-.19***	.02	.11***	.07***	.05***	.02	.20
SPSS PV5	.09***	.19***	.27***	-.18***	.03	.09***	.07***	.04**	.02	.20
SPSS PV6	.10***	.19***	.26***	-.18***	.03	.09***	.07***	.05***	.02	.19
SPSS PV7	.09***	.19***	.26***	-.20***	.03*	.09***	.08***	.05***	.03	.20
SPSS PV8	.08***	.19***	.26***	-.18***	.03*	.10***	.07***	.05***	.03	.19
SPSS PV9	.10***	.19***	.25***	-.20***	.02	.11***	.08***	.04**	.03*	.20
SPSS PV10	.09***	.19***	.27***	-.19***	.03*	.09***	.08***	.04**	.01	.20
SPSS PVort	.09***	.20***	.28***	-.20***	.03*	.10***	.08***	.05***	.02	.22

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

### ***Bir Bağımlı Değişkeni Çok Düzeyli Verisetleri ile Yordamayı Amaçlayan Çalışmalar***

Bu kısımda “TIMSS 2011 uygulamasında öğrenci düzeyi değişkenlerinden ailenin ödevin yapılmasını sağlaması, ailenin ödevi kontrol etmesi ve öğrencinin ödevi ayırdığı süre; öğretmen düzeyi değişkenlerinden öğretmenin ödevi kontrol etmesi, öğrencilere kontrol ettirmesi, ödevi tartışması, tamamlanmasını izlemesi, not için kullanması değişkenleri akıl yürütme puanlarını tahmin edebilir mi?” alt araştırma sorusu yanıtlanmaktadır. Hem öğrenciler hem de bu öğrencilerin sınıflarına giren öğretmen verisi kullanıldığı için çok düzeyli regresyon analizi yapılması gerekmektedir. Bu araştırma sorusu bu yapıdaki soruların yanıtlanmasında izlenecek yolu göstermeyi amaçlamaktadır. Sonuçlar Tablo 8’de verilmektedir.

Düzeyler arası korelasyon (Intraclass correlation) 0,32 bulunmuştur. Bu değer öğrenci verisinin bağımsız olmadığını, sınıflardaki öğrencilerin birbirine benzediğini ve bu sebeple çok düzeyli regresyon analizinin gerekli olduğunu göstermektedir. Çok düzeyli regresyon analizden elde edilen düzeyler arası korelasyon değeri göstermektedir ki, tüm olası değişkenler dikkate alınabilirse, öğrencilerin akıl yürütme becerilerindeki varyansın %32’si sınıflar arası varyanstan, %68’i ise öğrenciler arası varyanstan kaynaklanmaktadır. Bu çalışmada oluşturulan model ve kullanılan öğrenci düzeyindeki (düzey 1) değişkenler, öğrenciler arası varyansın %4’lük kısmını açıklayabilmektedir. Benzer şekilde sınıf düzeyindeki (düzey 2) öğretmenler ile ilgili değişkenler de sınıflar arası varyansın %7’lik kısmını açıklayabilmektedir. Bu değerler ortaya konulan modelin etkili olmadığını göstermektedir. Ancak bu çalışmanın amacı etkili model kurmaktan ziyade analizleri örneklendirmektir.

Öğrenci düzeyi değişkenlerinden ailenin ödevin yapılmasını sağlaması ve ailenin ödevi kontrol etmesi değişkenleri akıl yürütme puanlarını tahmin edebilmektedir. Ödevin yapılmasını sağlama ve akıl yürütme puanları arasında pozitif yönde bir ilişki varken, ailenin ödevi kontrol etmesi ve akıl yürütme

puanları arasında negatif yönde bir ilişki vardır. Sınıf seviyesindeki değişkenlerden öğretmenin ödevin tamamlanmasını izlemesi ile akıl yürütme arasında pozitif yönde bir ilişki bulunmuştur.

Tablo 8. Çokdüzeyli Regresyon Analizi Standart Katsayılar

Değişkenler	Katsayılar
<i>Düze-1</i>	
ailenin ödevin yapılmasını sağlaması	.17***
ailenin ödevi kontrol etmesi	-.19***
öğrencinin ödevi ayırdığı süre	-.03
<i>Düze-2</i>	
öğretmenin ödevi kontrol etmesi	-.04
öğretmenin öğrencilere ödevi kontrol ettirmesi	-.05
öğretmenin ödevi tartışması	.10
öğretmenin ödevin tamamlanmasını izlemesi	.16*
öğretmenin ödevi not için kullanması	.08
<i>Sınıflar Arası Açıklanan Varyans</i>	7%
<i>Sınıf İçi Açıklanan Varyans</i>	4%

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

## TARTIŞMA

Günümüzde geniş ölçekli test sonuçlarının birçok farklı ülkede eğitim politikaları ve yapılacak yatırımların belirlenmesinde, müfredat değişiklikleri ve eğitimde yenilikçi uygulamaların kullanımına karar verilme süreçlerinde önemli bir rol oynadığı bilinmektedir (Hamilton, 2003). Test sonuçları ayrıca çeşitli boyutlarda ülkelerarası karşılaştırma imkanı sağlamakta ve ülkelere kendi eğitim sisteminde yer alan çeşitli öğeler hakkında detaylı bilgiler de sunmaktadır. Politika belirlemede önemli rol oynaması sebebiyle, milyonlarca paydaşı ilgilendiren bu testlerden elde edilen verilerin hatasız ve doğru bir şekilde analiz edilmesi gerekmektedir. İlgili alan yazında PISA şoku (*PISA 'shock' phenomenon*) adıyla da (Wiseman, 2013) bilinen durum örneklerinde de görüldüğü üzere, geçmişte medyanın da etkisiyle geniş ölçekli test veri setlerinin yanlış yorumlanması, ilksel ve betimsel çıkarımlar üzerinden ülkelerin gerçekleştirdiği yanlış ve radikal politika uygulamaları gerçekleştirilmiştir. Özellikle Almanya'da 2000 yılında ilk kez dahil olduğu PISA uygulama sonuçlarının diğer ülke ortalamalarının altında kaldığının görülmesi sonrası gerçekleştirilen sert politika geçişleri (Waldow, 2009) ve Japonya'nın PISA 2000-2003 okuma ile 2003-2006 matematik alanındaki düşüşleri sonrası gerçekleştirilen politika değişiklikleri bu durumlara örnek olarak gösterilebilmektedir (Wiseman, 2013). Bu durum da test verilerinin derinlemesine ikincil olarak doğru teknik ve yazılımlar kullanılarak analizini zorunlu kılmaktadır (Wiseman, 2013). Bu çalışma kapsamında da, bu verilerin uygun istatistiksel analiz yöntemleriyle analiz edilmediği durumlarda elde edilebilecek hatalı sonuçlar, örneklem ağırlıkları ve olası değerlerin kullanımı bağlamında örneklendirilmiş, araştırmacılara doğru tekniklerin kullanılması konusunda yol gösterilmesi amaçlanmıştır.

Örneklem ağırlıkları ve olası değerlerin uygun kullanılmaması çeşitli değişkenlerine göre incelenen test sonuçlarında hatalı sonuçlara ulaşılabilmesine yol açabilmektedir. Örneğin, bu çalışmada görüldüğü üzere, uygun analiz yöntemleri kullanıldığında TIMSS 2015 Türkiye örnekleminde kızlar ve erkeklerin matematik ortalamaları arasında istatistiksel olarak anlamlı bir fark bulunmamaktadır. Ancak uygun analiz yöntemi kullanmadığında istatistiksel olarak anlamlı fark bulunduğu ve hatalı sonuçların ortaya çıktığı görülmektedir. Benzer şekilde, PISA 2015 uygulamasında fen dersindeki

disiplin iklimi, epistemolojik inançlar, sosyoekonomik statü, araştırma temelli fen öğretimi, araçsal motivasyon, feni sevme, fende özyeterlik, öğretmen merkezli fen öğretimi, öğrenci isteğine bağlı öğretmen desteği gibi bağımsız değişkenleri kullanarak öğrencilerin fen performans puanları bağımlı değişkenini yordayan çoklu regresyon analizi sonuçlarında da örneklem ağırlığı ve olası değerler dikkate alınmadan yapılan analiz sonuçlarına göre 11 veri setinin 8'i hatalı sonuçlar verdiği görülmüştür. Von Davier vd. (2009) ve Rutkowski vd. (2010)'nin Bulgaristan TIMSS 2007 uygulaması örneği üzerinden de vurguladığı üzere, örneklem ağırlıkları ve olası değerlerinin test verilerinin analizlerinde kullanımı doğru sonuçlara ulaşılması açısından büyük önem arz etmektedir.

İlgili alan yazın incelendiğinde, bu tür test verilerinin analizlerinde örneklem ağırlıkları ve olası değerlerin kullanımına ilişkin yeterli farkındalığın oluşmadığı görülmektedir. Bu iki özelliği sürece dahil eden doğru analizlerin yapıldığı çalışmalara rastlandığı gibi, puanlardan birini kullanan veya ortalamalarını alan hatalı çalışmaların da mevcut olduğu görülmektedir. Bu çalışmanın amacı doğru yöntemler hakkında bilgi vermek olduğu için doğru yöntemleri kullanmayan çalışmalar listelenmemiştir. Ancak, araştırmacılar yayınlara atıf yaparken veya sonuçları yorumlarken doğru teknikleri kullanan yayınları seçmeye özen göstermelidirler.

Türkiye'de bu verilerin nasıl analiz edileceği konusunda sistematik olarak açılan ders neredeyse yoktur. Sınırlı sayıda yapılan çalıştayların da ulaşabildiği araştırmacı sayısı azdır. Bu sınırlılıkları aşabilmek için bu çalışma hem olası değerler ve örneklem ağırlıkları hakkında farkındalık yaratarak araştırmacıları doğru analiz yöntemlerini kullanmaya yönlendirmek, hem de temel analizleri örneklendirerek araştırmacılar için yol göstermektir. Farklı veri setlerinde farklı örneklem ağırlıkları kullanıldığından dolayı, doğru örneklem ağırlığı kullanımları için araştırmacıların analizde kullanılan yazılımların özelliklerini ve alan yazında yer alan teknik raporları dikkatle incelemeleri de önemle önerilmektedir.

## KAYNAKÇA

- Addey, C., & Sellar, S. (2018). Why do countries participate in PISA? Understanding the role of international large-scale assessments in global education policy. In A. Verger, M. Novelli & H. Kosar-Altinyeken (Eds.), *Global education policy and international development: New agendas, issues and policies*, (pp. 97-118). New York, NY: Bloomsbury Publishing.
- Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., & Verger A. (2017). Forum discussion: The rise of international large-scale assessments and rationales for participation. *Compare*, 47(3), 434-452. doi:10.1080/03057925.2017.1301399
- Aydın, A., Selvitopu, A., & Kaya, M. (2018). Eğitime yapılan yatırımlar ve PISA 2015 sonuçları karşılaştırmalı bir inceleme. *İlköğretim Online*, 17(3), 1283-1301.
- Bialecki, I., Jakubowski, M., & Wiśniewski, J. (2017). Education policy in Poland: The impact of PISA (and other international studies). *European Journal of Education*, 52(2), 167-174.
- Carvalho, L. M. & Estela Costa, E. (2015) Seeing education with one's own eyes and through PISA lenses: considerations of the reception of PISA in European countries, *Discourse: Studies in the Cultural Politics of Education*, 36(5), 638-646. doi:10.1080/01596306.2013.871449
- Ertl, H. (2006). Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education*, 32(5), 619-634.
- Figazzolo, L. (2009). *Testing, ranking, reforming: Impact of PISA 2006 on the education policy debate*. Brussels: Education International.
- Froese-Germain, B. (2010). The OECD, PISA and the impacts on educational policy. *Canadian Teachers' Federation (NJ1)*. Retrieved from <http://eric.ed.gov/?id=ED532562>
- Gonzalez, E. J. (2012). Rescaling sampling weights and selecting mini-samples from large-scale assessment databases. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 5, 117-134.
- Gür, B. S., Celik, Z., & Özoğlu, M. (2012). Policy options for Turkey: A critique of the interpretation and utilization of PISA results in Turkey. *Journal of Education Policy*, 27(1), 1-21.
- Hamilton, L. (2003). Assessment as a policy tool. *Review of research in education*, 27(1), 25-68.

- International Association for the Evaluation of Educational Achievement (IEA) (2019). IDB Analyzer (version 4.0). Hamburg, Germany: IEA Hamburg.
- Landahl, J. (2018): De-scandalisation and international assessments: the reception of IEA surveys in Sweden during the 1970s. *Globalisation, Societies and Education*, 16(5), 566-576. doi:10.1080/14767724.2018.1531235
- LaRoche, S., & Foy, P. (2016). Sample design in TIMSS Advanced 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS Advanced 2015* (pp. 3.1–3.27). Erişim adresi <http://timssandpirls.bc.edu/publications/timss/2015-a-methods/chapter-3.html>
- LaRoche, S., & Foy, P. (2016). Sample implementation in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 5.1-5.175). Retrieved from <http://timss.bc.edu/publications/timss/2015-methods/chapter-5.html>
- Laukaityte, I., & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics-Theory and Methods*, 46(22), 11341-11357.
- Martens, K., & Niemann, D. (2010). Governance by comparison: How ratings & rankings impact national policy-making in education (No. 139). *TranState Working Paper*. Bremen: University of Bremen Collaborative Research Centre
- Milli Eğitim Bakanlığı (MEB). (2014). *TIMSS 2011 ulusal matematik ve fen raporu 8. sınıflar*. Ankara: T.C. Milli Eğitim Bakanlığı Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü.
- Milli Eğitim Bakanlığı (MEB). (2016a). *PISA 2015 ulusal raporu*. Ankara: T.C. Milli Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü.
- Milli Eğitim Bakanlığı (MEB). (2016b). *TIMSS 2015 ulusal matematik ve fen bilimleri ön raporu 4. ve 8. sınıflar*. Ankara: MEB: Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü.
- Milli Eğitim Bakanlığı (MEB). (2018). 2018 Liselere Geçiş Sistemi (LGS): Merkezi sınavla yerleşen öğrencilerin performansı. *Eğitim Analiz ve Değerlendirme Raporları Serisi (No. 3)*. Ankara: T.C. Milli Eğitim Bakanlığı.
- Michel, A. (2017). The contribution of PISA to the convergence of education policies in Europe. *European Journal of Education*, 52(2), 206-216.
- Monseur, C., & Adams, R. (2009). Plausible values: How to deal with their limitations. *Journal of Applied Measurement*, 10(3), 1-15.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Retrieved from <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide*. (7th ed.). Los Angeles, CA: Muthén and Muthén.
- Novoa, A. & Yariv-Mashal, T. (2003). Comparative research in education: A mode of governance or a historical journey? *Comparative Education*, 39(4), 423–438.
- The Organisation for Economic Co-operation and Development (OECD). (2016). *Skills matter: Further results from the survey of Adult Skills*. OECD Skills Studies. Paris: OECD Publishing. doi:10.1787/9789264258051-en.
- The Organisation for Economic Co-operation and Development (OECD). (2017). *PISA 2015 Technical Report*. Paris: OECD Publishing. Retrieved from <https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf>
- The Organisation for Economic Co-operation and Development (OECD). (2019). *PISA 2018 Assessment and Analytical Framework*. PISA. Paris: OECD Publishing. doi:10.1787/b25efab8-en.
- Pizmony-Levy, O. (2018). Compare globally, interpret locally: international assessments and news media in Israel. *Globalisation, Societies and Education*, 16(5), 577-595. doi:10.1080/14767724.2018.1531236
- Rubin, D. (1987). *Multiple imputation for nonresponse in sample surveys*. New York: John Wiley.
- Rust, K. (2013). Sampling, weighting, and variance estimation in international large-scale assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (1st ed., pp. 117–154). New York, NY: Chapman and Hall/CRC Press.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142-151.



- Steiner-Khamsi, G. & Waldow, F. (2018). PISA for scandalisation, PISA for projection: the use of international large-scale assessments in education policy making – an introduction. *Globalisation, Societies and Education*, 16(5), 557-565. doi:10.1080/14767724.2018.1531234
- Tiana Ferrer, A. (2017). PISA in Spain: Expectations, impact and debate. *European Journal of Education*, 52, 184-191.
- TEDMEM. (2016). *OECD yetişkin becerileri araştırması: Türkiye ile ilgili sonuçlar*. Ankara: Türk Eğitim Derneği Yayınları.
- Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI Monograph Series*, 2(1), 9-36.
- Waldow, F. (2009). What PISA did and did not do: Germany after the 'PISA-shock'. *European Educational Research Journal*, 8(3), 476-483.
- Wiseman, A. (2013). Policy responses to PISA in comparative perspective. In H.D. Meye, & A. Benavot (Eds.) *PISA, power, and policy: The emergence of global educational governance*. (pp.303–322). Oxford: Symposium Books
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3), 114-128.

**EK A. Birinci Araştırma Sorusu\_A**

Include file =

"C:\Users\exper\AppData\Roaming\IEA\IDBAnalyzerV4\bin\Data\Templates\SPSS\_Macros\JB\_PV.i  
easps".

```
JB_PV infile="D:\idb\TIMSS_2015.sav"/  
      cvar=IDCNTRY BSBG01 /  
      almvars=/  
      rootpv=BSMMAT0 /  
      tailpv=/  
      npv=5/  
      wgt=TOTWGT/  
      nrwt=150 /  
      rwgt=/  
      jkz=JKZONE/  
      jkr=JKREP/  
      jk2type=FULL/  
      nomiss=Y/  
      method=JRR/  
      kfac=0/  
      shrtcut=N/  
      viewcod=N/  
      ndec=2/  
      clean = Y/  
      strctry = N/  
      intavg = Y/  
      graphs=Y/  
      selcrit = /  
      selvar = /  
      outdir="D:\idb"/  
      outfile="PVMath_gender".
```

**EK B. Birinci Araştırma Sorusu\_B**

include file =

"C:\Users\exper\AppData\Roaming\IEA\IDBAnalyzerV4\bin\Data\Templates\SPSS\_Macros\JB\_PV.i  
easps".

```
JB_PV infile=" D:\idb\prgturp1.sav"/  
      cvar=CNTRYID C_Q02A /  
      almvars=/  
      rootpv=PVLIT /  
      tailpv=/  
      npv=10/  
      wgt=SPFWT0/  
      nrwt=80 /  
      rwgt=SPFWT/  
      jkz=/  
      jkr=/  
      jk2type=HALF/  
      nomiss=Y/  
      method=PIAAC/  
      kfac=0/  
      shrtcut=N/  
      viewcod=N/  
      ndec=2/  
      clean = Y/  
      strctry = N/  
      intavg = Y/  
      graphs=Y/  
      selcrit = /  
      selvar = /  
      outdir=" D:\idb"/  
      outfile="paidjoblook".
```

### EK C. İkinci Araştırma Sorusu

include file =

"C:\Users\Toshibanb\AppData\Roaming\IEA\IDBAnalyzerV4\bin\Data\Templates\SPSS\_Macros\JB\_RegGP.ieasps".

JB\_RegGP infile="C:\idb\PISA\_TUR2015.sav"/

```
cvar=CNTRYID /
convar=DISCLISCI EPIST ESCS IBTEACH INSTSCIE JOYSCIE SCIEEFF TDTEACH
TEACHSUP /
catvar=/
codings=/
refcats=/
ncats=/
PVRoots=/
PVTails=/
dvar0=/
rootpv=PV /
tailpv=SCIE /
npv=10/
wgt=W_FSTUWT/
nrwgt=80 /
rwt=W_FSTURWT/
jkz=/
jkr=/
jk2type=/
nomiss=Y/
method=BRR/
missing=listwise/
kfac=0.5/
shrtcut=N/
viewcod=N/
ndec=2/
clean = Y/
strctry = N/
viewprgs=Y/
viewlbl=Y/
qcstats=Y/
newout=Y/
intavg = Y/
selcrit = /
selvar = /
outdir="C:\idb"/
outfile="regression".
```

### **EK D. Üçüncü Araştırma Sorusu**

TITLE: this is an example of a two-level  
regression analysis

DATA: FILE = dataimputedlist.dat;  
!Create a file list;

TYPE = IMPUTATION;  
!Define that your data has multiple imputation;

VARIABLE:

NAMES = IDSCHOOL IDSTUD BSBG11C BSBG11D BSBM20B  
BTBM25CA BTBM25CB BTBM25CC BTBM25CD BTBM25CE  
REAPV WGTADJ1WGTFAC1 WGTADJ2WGTFAC2WGTADJ3WGTFAC3;

USEVARIABLES ARE IDSCHOOL BSBG11C BSBG11D BSBM20B  
BTBM25CA BTBM25CB BTBM25CC BTBM25CD BTBM25CE  
REAPV WGTADJ1WGTFAC1 WGTADJ2WGTFAC2WGTADJ3WGTFAC3;

CLUSTER = IDSCHOOL;  
!Define Cluster Variable here;

MISSING = ALL (9999);

WEIGHT = WGTADJ1WGTFAC1;  
BWEIGHT = WGTADJ2WGTFAC2WGTADJ3WGTFAC3;  
!Define Sample Weights Here;

WITHIN = BSBG11C BSBG11D BSBM20B;  
!Define Level1 variables here;

BETWEEN = BTBM25CA BTBM25CB BTBM25CC BTBM25CD BTBM25CE;  
!Define Level2 variables here;

ANALYSIS: TYPE = TWOLEVEL;  
!Define number of level here;

MODEL:

% WITHIN%  
REAPV on BSBG11C BSBG11D BSBM20B;

!Define Level1 relationships here;

% BETWEEN%  
REAPV on BTBM25CA BTBM25CB BTBM25CC BTBM25CD BTBM25CE;

!Define Level2 relationships here;

OUTPUT: STANDARDIZED;  
!For standardized coefficients;